

3 Monetarismus I: Geldmenge und Beschäftigung

Der neoklassische Monetarismus à la Friedman geht von der Annahme aus, dass die Ökonomie im Prinzip auf ein Vollbeschäftigungsgleichgewicht hin steuert. Gleichzeitig wird unterstellt, dass dieses Gleichgewicht ausschließlich durch reale Faktoren determiniert werde. Dies impliziert, dass Geld neutral ist, ein Schleier, der weggezogen werden muss, um die fundamentalen Gesetzmäßigkeiten der Ökonomie aufzudecken. Von diesem Stabilitätspostulat ausgehend, stellt Friedman fest, dass die monetären Impulse kurzfristig über die Veränderung des Nominaleinkommens reale Effekte auslösen, aber langfristig in die Veränderung des Preisniveaus ohne reale Wirkungen münden. Das heißt, die Phillipskurve ist langfristig vertikal. Friedman begründet dies zum einen mit der Idee der „natural rate of unemployment“ und zum anderen mit der neuformulierten Quantitätstheorie, die (mindestens langfristig) den Zusammenhang von Geldmenge und Preisniveau postuliert. Die monetaristische Logik geht von diesen beiden Ideen aus. Wirtschaftspolitisch laufen diese Argumente darauf hinaus, dass die Geldpolitik den potentialorientierten Geldmengenregeln folgen soll und das Problem der Arbeitslosigkeit allein durch angebotsorientierte Politik gelöst werden kann. Damit wurde das neoklassisch-monetaristische „assignment“ auf den Plan gerufen.

3-1 Die natürliche Arbeitslosenquote

Wir konzentrieren uns zunächst auf das Konzept der natürlichen Arbeitslosenquote (NRU).¹⁵ Friedman definiert die natürliche Arbeitslosenquote so:

„At any moment of time, there is some level of unemployment which has the property that it is consistent with equilibrium in the structure of real wage rates (...) The ‚natural rate of unemployment‘, in other words, is the level that would be ground out by the Walrasian system of general equilibrium equations, provided there is imbedded in them the actual structural characteristics, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.“ (Friedman 1968: 8).

¹⁵ Phelps (1967, 1970) entwickelte gleichzeitig die Hypothese der natürlichen Arbeitslosenquote. Für die kritische Auseinandersetzung mit der NRU vergleiche Hahn (1995) und Tobin (1995).

Die NRU ist konzipiert als ein eindeutiges langfristiges Gleichgewicht. Die Besonderheit dieser Definition besteht darin, dass die NRU ausschließlich durch reale Bedingungen wie Präferenzen und technisches Wissen determiniert wird. Wie der Terminus „natural“ – wie in der Verbindung mit dem Zinssatz bei Wicksell – suggeriert, wird die gleichgewichtige Beschäftigung als unabhängig von monetären Faktoren konzipiert. Die NRU als eine Angebotskategorie bleibt von der makroökonomischen Politik unberührt. Die Durchsetzung der NRU-Hypothese impliziert somit, dass der Nachfragepolitik die Grundlage entzogen wird.

Dem Konzept der NRU liegt die Logik der Vollbeschäftigung zugrunde.¹⁶ Nach dieser Logik funktioniert der Arbeitsmarkt ebenso wie beliebige andere Märkte walrasianisch. Das heißt, der Reallohn koordiniert das Angebot an und die Nachfrage nach Arbeit und führt zu einem markträumenden Gleichgewicht. Entsprechend findet sich der Arbeitsmarkt stets zu einem Zustand der Vollbeschäftigung, solange der Preismechanismus nicht von außen blockiert wird.

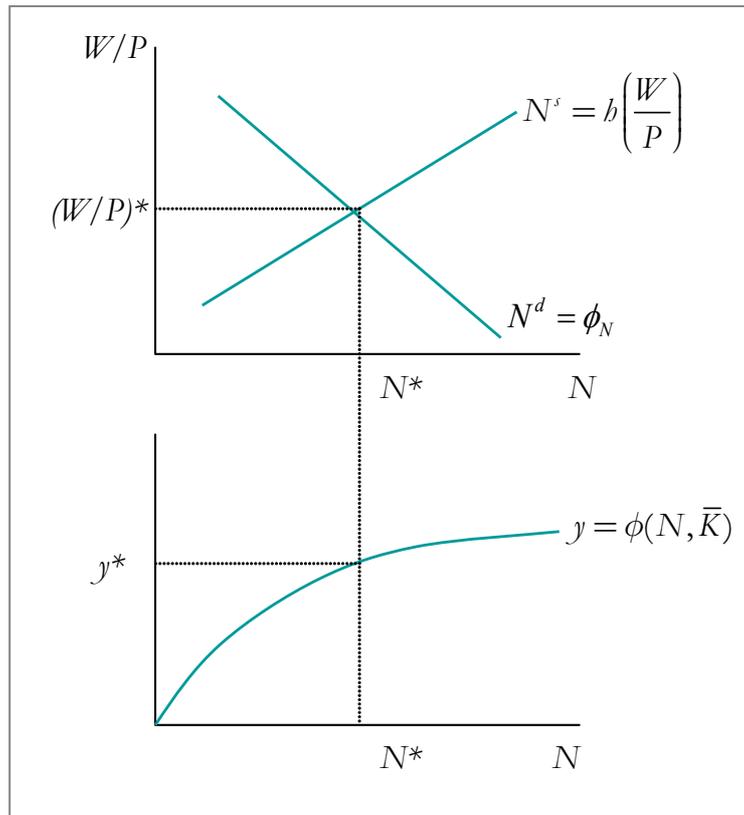
Die Nachfragefunktion nach Arbeit leitet sich aus einer Produktionsfunktion ab, die die Produktionsmöglichkeiten für die Unternehmer angibt. Jeder gewinnmaximierende Unternehmer setzt die Produktionsfaktoren in dem Verhältnis ein, in dem ihre (Wert-)Grenzproduktivitäten ihren Preisen entsprechen. Bei gegebenem Kapitalstock – dies ist die gewöhnliche, kurzfristige Annahme – und substitutionaler Produktionsfunktion fällt die Grenzproduktivität der Arbeit mit steigender Beschäftigung. Da das gewinnmaximierende Unternehmen die Arbeit mit ihren (Wert-)Grenzprodukt entlohnt, lässt sich die fallende Grenzproduktionsfunktion der Arbeit zugleich als Arbeitsnachfragefunktion betrachten. Die Arbeitsnachfrage des Unternehmens ist somit negativ vom Reallohn abhängig.

Die Angebotsfunktion an Arbeit leitet sich aus dem mikroökonomischen Nutzenkalkül der Individuen ab. Die nutzenmaximierenden Individuen vergleichen den Nutzen der Arbeitszeit mit dem Nutzen der Freizeit. Ein steigender Lohn bedeutet, dass der Nutzen einer Stunde Arbeit im Vergleich zu einer Stunde Freizeit zunimmt. Die Individuen reagieren darauf mit einem höheren Angebot von Arbeit. Das heißt, das Arbeitsangebot korreliert positiv mit dem Reallohn.

Der Reallohn koordiniert damit das Angebot und die Nachfrage der Arbeit. Aus diesem Blickwinkel hat Friedman beurteilt, dass ein „basic defect“ von Phillips' Analyse in der ungenügenden Unterscheidung zwischen realen und nominalen Größen begründet liege.

¹⁶ Nach Dixon (1995) hat Patinkin (1965: 127-130) erstmals das Vollbeschäftigungsmodell des Arbeitsmarktes explizit präsentiert. Danach wurde sein Modell zum Standardmodell der Lehrbücher.

Abbildung 3.1 Das Arbeitsmarktmodell der Vollbeschäftigung



Die Abbildung 3.1 illustriert die Funktionsweise des Arbeitsmarktes.¹⁷ Es gibt einen Reallohn, $(W/P)^*$, zu dem die Angebotsmenge gleich der nachgefragten Menge an Arbeit ist. Diesem (gleichgewichtigen) Reallohn korrespondiert das gleichgewichtige Beschäftigungsniveau N^* . Ein Reallohn über dem vollbeschäftigungskonformen Niveau führt zum Angebotsüberschuss, der wiederum sinkenden Reallohn bewirkt; im Fall des Nachfrageüberschuss ist es umgekehrt. Mithin wird argumentiert, dass der Arbeitsmarkt stets über einen flexiblen Reallohn geräumt wird. Entsprechend wird die Arbeitslosigkeit als Resultat – aus welchen Gründen immer – eines falschen Preises, nämlich eines zu hohen Reallohnes, diagnostiziert.

¹⁷ Die Nachfragekurve nach Arbeit ist aus einer Produktionsfunktion $y = \phi(N, \bar{K})$ abgeleitet, d.h. $N^d = \phi_N$, wobei $(W/P) = \phi_N$.

Die Abbildung 3.1 verdeutlicht diese Aspekte. Sie schließt eine Produktionsfunktion ein, wobei der gleichgewichtige Output y^* mit der gleichgewichtigen Beschäftigung N^* korrespondiert. Die aggregierte Angebotskurve verläuft dann vertikal bei diesem Beschäftigungsniveau. Ersichtlich ist, dass Beschäftigung und Reallohn sowie der Output innerhalb des Arbeitsmarktes determiniert werden. Aus diesem Grund erklären Felderer/Homburg (2003: 85ff.), dass der Arbeitsmarkt im neoklassischen Modell eine strategische Rolle spielt: Auf dem Arbeitsmarkt entscheidet sich alles. Das heißt, dass Reallohn und Beschäftigung allein auf dem Arbeitsmarkt durch individuelle Entscheidungen bestimmt werden.

Welche Bedeutung hat das Konzept der natürlichen Arbeitslosenquote? Friedman hat mit der NRU das langfristige Gleichgewicht in Form seiner zeitgenössischen Arbeitsmarkttheorie, nämlich des Suchmodells, dargestellt (Friedman 1977). Die Abweichung vom Vollbeschäftigungsgleichgewicht ist eine Folge der unvermeidbaren Marktunvollkommenheit und mithin als „natürliche“ Rate identifiziert. Dieser Umstand läuft darauf hin, dass die von der Arbeitsmarktstatistik ausgewiesene Arbeitslosigkeit im Wesentlichen freiwilliger Natur sei. Noch wichtiger ist die Rolle der natürlichen Arbeitslosenquote für die Geldpolitik. Friedman beschreibt dies so:

„I introduce the concept in a section titled ‚What Monetary Policy Cannot Do‘. It was part of an explanation of why, in my opinion, the monetary authority cannot adopt ‚a target for employment or unemployment (...); be tight when unemployment is less than the target; be easy when unemployment is higher than the target‘“ (Friedman, Wall Street Journal 24.9.1996, zitiert nach Espinosa-Vega/Russell 1997: 13).

Die NRU markiert damit einen Anker für die Geldpolitik – eine Idee, die in dem NAIRU (Non Accelerating Inflation Rate of Unemployment) Konzept ausgeformt wurde. Die Integration der NRU in die Phillipskurve leitet Friedman zur Formulierung einer (langfristig) vertikalen Phillipskurve und damit zur Behauptung, dass ein Versuch, auf der Phillipskurve zu reiten, in akzelerierende Inflationsraten münde. „The only way unemployment can be kept below the ‚natural rate‘ is by an ever-accelerating inflation, which always keeps current inflation ahead of anticipated inflation“ (Friedman 1975: 74). Wenden wir uns nun der monetären Analyse von Friedman zu.

3-2 Realkasseneffekt und Neutralität des Geldes

3-2-1 Neue Formulierung der Quantitätstheorie

Während die neoklassische Theorie die realökonomische Analyse mit dem walrasianischen Modell abdeckt, findet sich eine Ausformung ihrer monetären Analyse in der Quantitätstheorie. Wenn y für das reale Sozialprodukt, P für das Preisniveau, V für die Umlaufgeschwindigkeit und M für die nominale Geldmenge steht, lautet die Quantitätsgleichung

$$M \cdot V = P \cdot y \quad (3.1)$$

Die rechte Seite der Gleichung repräsentiert das nominale Sozialprodukt der Ökonomie und die linke Seite die nominale Geldmenge. Die Frage ist, wie die Ursache-Wirkungs-Zusammenhänge dieser an sich tautologischen Gleichung begründet werden. Die Quantitätstheorie postuliert eine proportionale Relation von Geldmenge und Preisniveau. Das Herzstück der Quantitätstheorie lautet also: $P = f(M)$. Für diese Relation sind drei Annahmen zentral:¹⁸

- Die Geldmenge wird exogen gesetzt. Damit wird eine Omnipotenz der Zentralbank in dem Sinne postuliert, dass sie die nominale Geldmenge beliebig festlegen kann.
- Das reale Sozialprodukt bzw. das reale Einkommen wird durch die Realsphäre bestimmt und ist unabhängig von monetären Prozessen.
- Die Wirtschaftssubjekte wünschen unter gegebenen Umständen eine bestimmte reale Geldmenge zu halten. Entsprechend wird eine stabile Umlaufgeschwindigkeit des Geldes – bzw. ein gegebener Kassenshaltungskoeffizient nach der Cambridge-Gleichung – angenommen.

Auf Basis dieser Annahmen leitet die Quantitätstheorie einen Marktprozess von der Geldmenge zum Preisniveau her. Eine Veränderung der nominellen Geldmenge schlägt sich in einer proportionalen Veränderung des Preisniveaus nieder, ohne Auswirkungen auf Output und Beschäftigung. Das reziproke Preisniveau bildet den Preis des Geldes. Die Kernaussagen der Quantitätstheorie lauten somit:

¹⁸ Dazu vgl. Heine/Herr (2003: 269ff.). Zur Exposition der Quantitätstheorie siehe Friedman (1974: 1ff.). Zur dogmenhistorischen Darstellung vgl. Claassen (1980) und Heise (1991).

- Erstens dass Geld (langfristig) neutral sei. Dieser Umstand manifestiert sich in der Dichotomie zwischen Geld- und Gütersphäre, in dem Sinne, dass beide ihren eigenen Gesetzmäßigkeiten folgen: Die Geldsphäre nach der Quantitätstheorie und die Gütersphäre nach dem walrasianischen Gesetz.
- Das Neutralitätspostulat impliziert weiterhin, dass die Inflation ihren Grund in der Expansion der Geldmenge habe. Friedman bringt dies so zum Ausdruck, dass die Inflation „always and everywhere“ ein monetäres Phänomen sei.

Die Geschichte der Quantitätstheorie reicht sehr viel weiter zurück als nur bis zum neoklassischen Monetarismus. Da jedoch das Verhalten der Wirtschaftssubjekte in den alten Ausformulierungen nicht explizit begründet wurde, hat die Neuformulierung der Quantitätstheorie durch Friedman (1970b) an diesem Punkt angesetzt. Der Ausgangspunkt seiner Neuformulierung der Quantitätstheorie war, dass die Quantitätstheorie in erster Linie eine Theorie der Geldnachfrage sei. Der Fokus wurde dabei auf die Stabilität der Geldnachfrage gelegt. Da das Geldangebot als exogen fixiert postuliert wird, bildet sie eine Voraussetzung für den Kausalnexus von Geldangebot auf Geldnachfrage. Dieser Punkt erklärt, warum das monetaristische Modell den Akzent auf eine Fundierung stabiler Geldnachfrage setzt (vgl. Issing 2001: Chap.1). Sonst wäre die Verbindung von Geldmenge und Preisniveau nicht zu halten.

Friedman hat, der keynesianischen Methodik folgend,¹⁹ versucht, die Geldhaltung mit den nichtpekuniären Erträgen auf die Geldhaltung zu begründen: Geld wird gehalten, weil es die Transaktionskosten senkt und Sicherheit erzeugt. Die Geldnachfrage hängt von drei Hauptfaktoren ab: Erstens vom Gesamtvermögen, welches als Budgetrestriktion die maximal zu haltende Geldmenge bestimmt; zweitens von den Erträgen auf Geld im Vergleich zu anderen finanziellen und realen Vermögensobjekten; drittens von den Präferenzen der Wirtschaftssubjekte. Das Gesamtvermögen wird in verschiedenen Vermögensformen in Abhängigkeit von deren relativen Erträgen gehalten,

¹⁹ Diese Formalie darf nicht über einen fundamentalen Unterschied hinweg täuschen: Das zentrale Argument für die Geldhaltung in der keynesianischen Ökonomie ist die „Annehmlichkeit und Sicherheit“ eines gehaltenen Geldvolumens, dessen nichtpekuniäre Erträge sich in der Liquiditätsprämie manifestieren. Riese (1986: 24) betont, dass die Liquiditätsprämie bei Keynes somit in den Kontext des Kalküls des Vermögensmarkt gestellt ist, während die Nequantitätstheorie die nichtpekuniären Erträge lediglich an einem Interesse am Konsum von Gütern und Dienstleistungen, also in einer auf den Gütermarkt projizierten Geldhaltung der Haushalte, sieht.

sodass deren Grenzerträge im Gleichgewicht gleich werden.²⁰ Die (makroökonomische) Geldnachfragefunktion lässt sich ausdrücken als:²¹

$$\frac{M^d}{P} = f(y; i_B, i_E, \pi^e; g) \quad (3.2)$$

wobei y das permanente Einkommen als Proxy für das Gesamtvermögen fungiert, i_B bzw. i_E als (erwartete) Ertragsrate von Obligationen bzw. Aktien die finanziellen Vermögensformen repräsentieren, während π^e für die erwartete Inflationsrate und g für die Präferenz der Wirtschaftssubjekte stehen. Es gilt:

$$\frac{\partial f}{\partial y} > 0; \quad \frac{\partial f}{\partial i_B} < 0; \quad \frac{\partial f}{\partial i_E} < 0; \quad \frac{\partial y}{\partial \pi^e} < 0; \quad \frac{\partial f}{\partial g} > 0$$

Die Geldnachfragefunktion (3.2) ist so konstruiert, dass sie als eine Anwendung von Friedmans permanenter Einkommenstheorie der Konsumtion angesehen werden kann und dass die erwartete Inflationsrate als ein Hauptargument eingeführt wird. Ein entscheidender Punkt von (3.2) besteht allerdings im Postulat, dass die Wirtschaftssubjekte eine reale Kasse nachfragen.

„The quantity theory takes for granted that what ultimately matters to holders of money is the real quantity rather than the nominal quantity they hold and that there is a fairly definite real quantity of money that people wish to hold under any given circumstances.” (Friedman 1974: 2).

Die Abhängigkeit der Geldnachfrage von realen Größen drückt sich so aus, dass die Geldnachfrage bezüglich des Preisniveaus homogen im ersten Grad ist. Das Preisniveau wird somit nicht mehr eine unabhängige Variable der Geldnachfragefunktion, sondern bloß Multiplikator der Nachfrage nach realer Kasse M^d/P angesehen (Heine/Herr 2003: 280). Mithin wird eine stabile Geldnachfrage behauptet: Die Geldnachfrage hängt von realen Größen ab, die im Wesentlichen stabil sind.

²⁰ Friedman (1970b) betont, dass seine Geldnachfragefunktion im Gegensatz zur keynesianischen verschiedene Vermögensformen als unabhängige Variablen einführt, um die Portfolioentscheidungen der Wirtschaftssubjekte zu berücksichtigen.

²¹ Heine/Herr (2003: 279) weisen jedoch darauf hin, dass Friedman keine strikte Verbindung zwischen der mikroökonomischen Analyse der Geldnachfrage und der makroökonomische Geldnachfragefunktion liefert.

Bezogen auf das Geldangebot folgt Friedman dem traditionellen Gedanken: Das Geldangebot wird exogen von der Zentralbank gesetzt. Friedman versinnbildlicht diese Unterstellung mit einem Banknoten abwerfenden Hubschrauber.²² Entsprechend wird die Zentralbank als monetäre Autorität, also als omnipotenter Träger des monetären Prozesses stilisiert. Da $M^s = \bar{M}$ ist, ergibt sich die Gleichgewichtsbedingung auf dem Geldmarkt als

$$\frac{\bar{M}}{P} = f(y; i_B, i_E, \pi^e; u) \quad (3.3)$$

Die Gleichung (3.3) ersetzt den Kassenhaltungskoeffizienten der Cambridge-Version der Quantitätstheorie durch eine stabile Realkassenhaltung. Aber der Kern der alten Quantitätstheorie bleibt gewahrt: Eine Erhöhung der Geldmenge durch die Zentralbank führt zu einer proportionalen Erhöhung des Preisniveaus. Geld(politik) ist somit neutral in Bezug auf die Realökonomie. Issing bewertet diese (langfristige) Neutralität des Geldes als allgemeinen Konsensus in der akademischen Literatur.

„It appears consensual that, in equilibrium, monetary policy is ‚neutral‘; eventually, a monetary policy shock is reflected, *ceteris paribus*, into a permanent change of the price level and no permanent changes in real variables (...) In the long run, there is therefore a clear dichotomy between the real and nominal variables in the economy.“ (Issing 2001: 8).

Es bleibt noch eine Frage offen – nämlich die Frage nach dem Marktprozess, der von einem Gleichgewichtszustand zum anderen führt. Es handelt sich um den so genannten Transmissionsmechanismus. Friedman erfasst diese Frage mit dem Realkasseneffekt.²³

Nehmen wir an, dass sich eine Ökonomie in einem Gleichgewichtszustand befindet und dass die Zentralbank die nominelle Geldmenge verdoppelt, die dann auf dem erhöhten Niveau verharren soll. Dadurch wird bei zunächst unverändertem Preisniveau die reale Geldmenge ebenfalls verdoppelt. Es ergibt sich bei der Vermögenshaltung des Publikums ein Ungleichgewicht zwischen gewünschter und tatsächlicher Geldmenge. Die reale Geldmenge, über die das Publikum verfügt, wird größer als das gewünschte (daher optimale) Niveau. An diesem Punkt tritt der Realkasseneffekt in Kraft. Das Publikum beginnt die Kasse durch Mehrausgabe (Pigou-Effekt) bzw. durch Umstrukturi-

²² Wir diskutieren später ausführlich die Konsequenzen der Annahme eines exogenen Geldangebots (siehe Kapitel 7).

²³ Für den dogmengeschichtlichen Überblick vgl. Cassens (1997).

rierung des Portfolios (Keynes-Effekt) abzubauen.²⁴ Dies führt zu einer Überschussnachfrage auf dem Gütermarkt, da das Güterangebot der Annahme nach realökonomisch gegeben ist. Die Überschussnachfrage lässt das Preisniveau steigen, und es kommt zu einem inflationären Prozess, wohl zunächst auch zu einer Erhöhung des Outputs.

Die anfänglich überschüssige Realkasse wird durch das ansteigende Preisniveau immer mehr reduziert. Im neuen Gleichgewicht geht die Inflation wieder auf Null, das Preisniveau hat sich verdoppelt, und die reale Geldmenge in Händen des Publikums hat wieder ihr altes optimales Niveau erreicht. Das heißt, der monetäre Impuls induziert allein die proportionale Veränderung des Preisniveaus, während andere realökonomische Größen wie Output und Beschäftigung nach einer vorübergehenden Störung unverändert bleiben. Das bedeutet wiederum, dass die ökonomische Entwicklung langfristig durch reale Faktoren bestimmt wird, während das Geld kurzfristig (störende) Einflüsse auf die Realökonomie haben kann.

Der Realkasseneffekt bildet die Steuerungsgröße, die die Ökonomie auf das Gleichgewicht führt. Er steuert die Geldhaltung des Publikums dergestalt, dass die neoklassischen Allokationsbedingungen des Gütermarktes wieder hergestellt werden. Die Verbindung zwischen Geldmenge und Preisniveau und damit das Neutralitätspostulat der Quantitätstheorie wird über den Realkasseneffekt bestätigt. Entscheidend ist dabei, dass der Realkasseneffekt die Geldsphäre sich an die Gütersphäre anpassen lässt und damit die erstere zum Anhängsel der letzteren macht.

An dieser Stelle soll jedoch auf den eigentlichen Schwachpunkt dieses Gedankengangs hingewiesen sein: Die Vorgabe einer exogenen Geldmenge entspricht lediglich der tauschwirtschaftlichen Logik – sie widerspricht jedoch der Funktionsweise der Geldwirtschaften. In der Geldwirtschaft entsteht Geld endogen über die Kreditbeziehung zwischen Zentralbank und Geschäftsbanken (dazu Kapitel 7). Ist die Geldmenge endogen, wird dem Realkasseneffekt die Grundlage entzogen: Kein Pigou-Effekt, weil Geld kein Reinvermögen ist (sprich: „Innengeld“); kein Keynes-Effekt, weil das über-

²⁴ Im Lehrbuch finden sich zwei Arten des Realkasseneffekts: der Pigou- und der Keynes-Effekt. Der Pigou-Effekt setzt voraus, dass der Realkasse Reinvermögen gegenübersteht. Da das Vermögen ein Argument der Konsumnachfragefunktion ist, führt eine Erhöhung der Realkasse direkt zur Konsumnachfrage. Der Keynes-Effekt hingegen erfolgt über den Geldmarkt. Eine erhöhte Realkasse führt das Publikum zur Umschichtung seines Vermögens, sodass die Nachfrage nach Bonds steigt. Der Kurs der Anleihen steigt, die Zinsen sinken und die Investitionsnachfrage wird angeregt. Den beiden Effekten ist gemeinsam, dass eine erhöhte Realkasse letztlich in eine erhöhte Nachfrage mündet. Startet das Argument also bei Vollbeschäftigung, so muss eine Erhöhung der Realkasse eine Überschussnachfrage induzieren.

schüssige Geld nicht in Kreditexpansion, sondern in seiner Vernichtung durch Abbau der Verbindlichkeiten gegenüber der Zentralbank resultieren kann. Zudem zeigt Betz (2001: 59f.) auf, dass der Geldmarkt bei Abwesenheit von Nominalillusion über das Preisniveau nicht geräumt werden kann. Endogene Geldmengenveränderung als Reflex der Kreditentscheidungen heißt einerseits, dass die Wirtschaftssubjekte mehr Kredite nachfragen, wenn das Preisniveau steigt; andererseits führt ein höheres Preisniveau zu einem höheren Kreditvolumen, das wiederum ein höheres nominales Geldangebot erzeugt. Daraus ergibt sich, dass bei einem höheren (bzw. niedrigeren) Preisniveau die reale Geldmenge konstant bleibt.

3-2-2 Monetäre Theorie des Nominaleinkommens

Die Neoquantitätstheorie hält am Neutralitätspostulat des Geldes fest, schließt jedoch kurzfristige Wirkungen der Geldpolitik nicht aus. Vor diesem Hintergrund stellt Friedman die Frage, wie der monetäre Impuls auf einen Inflations- und Realeffekt verteilt werden kann.

Friedman diskutiert diese Frage anhand des so genannten Common Model, das den Rahmen des IS-LM-Modells verwendet, um die unterschiedliche Lösungsansätze der neoklassischen und keynesianischen Ökonomie sowie seinen „third way“ zu präsentieren. In allgemeiner Form besteht das Common Model aus zwei Gleichungen:

$$\bar{M} = P \cdot L(i, y) \quad (3.4)$$

$$g(i) = y - c(i, y) \quad (3.5)$$

wobei P das Preisniveau, y das Realeinkommen, i der Zinssatz ist und L die Liquiditäts-, c die Konsum-, g die Investitionsfunktion bezeichnet. Zu beachten ist dabei, dass das Geldangebot als exogen fixiert angenommen wird, so dass gilt $M^s = \bar{M}$. Das Modell weist somit mit zwei Gleichungen und drei Variablen einen Freiheitsgrad auf, wobei (3.4) die Gleichgewichtsbedingung der Geldsphäre (bzw. LM-Kurve) darstellt, während (3.5) das Gütermarktgleichgewicht wiedergibt (bzw. IS-

Kurve).²⁵ Die paradigmatische Differenz zeigt sich in der Art und Weise, wie der Freiheitsgrad besetzt wird.

Ein Punkt der Gleichung (3.4) ist hervorzuheben. Ihr liegt das Homogenitätspostulat zugrunde, durch das die keynesianische Interdependenz zwischen Geldmenge und Zinssatz wegdefiniert wird. Die Realkasse passt sich über P an L an. Der Zinssatz als monetäre Kategorie ist dadurch unter den Realkasseneffekt subsumiert. Wie Riese (1986: 12ff.) betont, verbleibt das Common Model mit dem exogen fixierten Geldangebot und dem Homogenitätspostulat im Wesentlichen im Rahmen der neoklassischen monetaristischen Ökonomie.

Die einfache Quantitätstheorie schließt das System durch die exogene Fassung des Realeinkommens, sodass $y = \bar{y}$. Das Einkommen ergibt sich aus der mikrofundierten Allokation vorgegebener Ressourcen: Es bildet sich aus einer mit relativen Preisen bewerteten Erstausrüstung, und diese mikroökonomische Bestimmung drückt sich makroökonomisch als vorgegebenes Einkommen aus. Hinter dieser Annahme steht die Vorstellung, dass der Arbeitsmarkt im Vollbeschäftigungsgleichgewicht ist. Unter diesen Bedingungen bestimmt sich nun der gleichgewichtige Zinssatz gütermarktbezogen, während das Preisniveau, der quantitätstheoretischen Logik entsprechend, durch die exogen von den monetären Autoritäten bestimmte Geldmenge bestimmt wird. In dieser Lösung reflektiert sich die klassische Dichotomie zwischen Geld- und Gütersphäre.

Der Lösung der einfachen keynesianischen Ökonomie – darunter wird das IS-LM-Modell verstanden – unterliegt die Annahme eines konstanten Preisniveaus. Mithin wird eine simultane Bestimmung von Zinssatz und Einkommensbildung deduziert, wobei der Zinssatz durch die Liquiditätspräferenz der Wirtschaftssubjekte determiniert wird und das Einkommen die Beschäftigungshöhe angibt. Die Loslösung des Einkommens von den vorgegebenen Ressourcen weist die keynesianische Ökonomie als Beschäftigungstheorie aus.

Friedman meint, dass beide makroökonomischen Modelle, was die Erklärung der kurzfristigen Schwankungen betrifft, Defizite aufweisen. Er stellt fest, dass keines der beiden Modelle eine Aussage über die Faktoren enthält, die determinieren, wie sich ein monetärer Impuls – in der kurzen Frist – auf Preis- und Mengeneffekte aufteilt. Die einfache Quantitätstheorie unterstellt, dass der

²⁵ Das von Friedman (1974) präsentierte Modell besteht aus sechs Gleichungen, die sieben Variablen aufweisen. Dabei stellen drei Gleichungen die Gleichgewichtskonstellation für die Geldsphäre, die anderen drei Gleichungen die Gleichgewichtskonstellation für den Gütermarkt dar, so dass sich sie vereinfachend in die oben dargestellten zwei Modellgleichungen zusammenfassen lassen. Die Darstellung hier bezieht sich auf Riese (1986; 1989).

monetäre Impuls durch das Preisniveau absorbiert wird und der Output immer auf seinem Gleichgewichtsniveau bleibt. Die einfache keynesianische Ökonomie nimmt dagegen an, dass ein monetärer Impuls immer dann zu Mengeneffekten führt, wenn Unterbeschäftigung vorliegt; bei Vollbeschäftigung sind die quantitätstheoretischen Zusammenhänge jedoch wirksam.

Aus diesen Gründen²⁶ stellt Friedman eine „monetary theory of nominal income“ als „third way“ der Lösung vor, die allein auf die monetäre Wirkung auf das Nominaleinkommen zentriert und auf eine explizite Zerlegung des Nominaleinkommens in das reale Einkommen und das Preisniveau verzichtet. Die konstitutiven Elemente seiner Lösung sind dabei

„(a) A unit elasticity of the demand for money with respect to real income; (b) A nominal market interest rate equal to the anticipated real rate plus the anticipated rate of change in prices, kept at that level by speculators with firmly held anticipations; (c) A difference between the anticipated real interest rate and the real secular rate of growth determined outside the system; (d) Full and instantaneous adjustment of the amount of money demanded to the amount supplied“ (Friedman 1974: 42f.).

Aufgrund von (b) und (c) postuliert Friedman, dass kurzfristig sowohl der Real- als auch der Nominalzinssatz konstant sind – eine Annahme, durch die das System geschlossen wird.²⁷ Diese Annahme $i = \bar{i}$ spielt die gleiche Rolle wie \bar{y} und \bar{P} für die beiden oben genannten Modelle. Unter diesen Bedingungen lässt sich die Gleichung (3.4) so umformulieren:

$$\bar{M} = Y \cdot L(\bar{i}) \quad (3.6)$$

bzw.

$$Y = V(\bar{i}) \cdot \bar{M} \quad (3.7)$$

²⁶ In diesem Kontext schreibt Friedman: “The key need to remedy the defects common to all the models (...) is a theory that will explain (a) the short-run division of a change in nominal income between prices and output, (b) the short-run adjustment of nominal income to a change in autonomous variables and (c) the transition between this short-run situation and a long-run equilibrium” (Friedman 1974: 48).

²⁷ Nach der Bedingung (b) ist $i = r^e + \pi^e$ mit r als realem Zinssatz, π als Inflationsrate. Da $\pi^e = \hat{Y}^e - \hat{y}^e$ ist, wird diese Gleichung zu $i = (r^e - \hat{y}^e) + \hat{Y}^e = k_0 + \hat{Y}^e$, wobei k_0 eine Konstante ist. Somit $\hat{i} = \bar{i}$. Frisch (1983: 110) weist darauf hin, dass Friedman zögert, diese Annahme des konstanten Zinssatzes auf die Gütersphäre anzuwenden. Denn diese impliziert einen konstanten realen Zinssatz und schließt damit eine zentrale ökonomische Determinante der Ersparnis und Investition aus. Friedman beschränkt daher seine Analyse auf die Formulierung einer monetären Theorie des Einkommens, betrachtet den „saving-investment sector as unfinished business“ (Friedman 1974: 40).

wobei die Umlaufgeschwindigkeit V eine Funktion des Zinssatzes ist.

Die Gleichungen (3.6) und (3.7) stellen den ersten Schluss dar, dass eine Änderung des Nominaleinkommens stets das Resultat einer Änderung des Geldangebots ist. Man kann auch diesen Schluss in dynamischer Form beschreiben. Der Grundgedanke ist dabei, dass im langfristigen Gleichgewicht alle Erwartungen realisiert werden, während kurzfristig tatsächliche und erwartete Werte der Variablen voneinander abweichen können, wodurch ein Anpassungsprozess ausgelöst wird.

Die Totalableitung der Gleichung (3.7) nach der Zeit gibt dem Modell einen dynamischen Charakter. Unter Berücksichtigung der relativen Veränderung der Umlaufgeschwindigkeit $s = (1/V)(dV/di)$ und $\hat{i} = \hat{Y}^e$ (siehe Fn. 28) sowie der Spezifizierung des Anpassungsprozesses unter der Annahme adaptiver Erwartungen in Form von $\frac{d}{dt}(\hat{Y}^e) = \beta(\hat{Y} - \hat{Y}^e)$ erhält man

$$\hat{Y} = \hat{Y}^e + \frac{1}{1 - \beta s} (\hat{M} - \hat{Y}^e) \quad (3.8)$$

wobei die Stabilitätsannahme $(1 - \beta s) > 0$ ist. Die Gleichung (3.8) wird als eine kurzfristige monetäre Theorie des Nominaleinkommens betrachtet (Frisch 1983: 112). Wenn kurzfristig die Wachstumsrate der Geldmenge die erwartete Wachstumsrate des Nominaleinkommens übersteigt, d.h. $\hat{M} > \hat{Y}^e$, liegt das realisierte Wachstum über dem erwarteten Nominaleinkommen ($\hat{Y} > \hat{Y}^e$) und vice versa. Im (langfristigen) Gleichgewicht gilt jedoch $\hat{Y} = \hat{Y}^e = \hat{M}$.

Friedman versucht nun zu erklären, in welchem Ausmaß eine Änderung des Nominaleinkommens in einem Preis- und in einem Outputeffekt mündet. Er nennt zwei grundlegende Einflussfaktoren: zum einen die Erwartungen über das Verhalten des Preisniveaus und zum anderen das gegenwärtige Niveau des Outputs bzw. der Beschäftigung im Vergleich zum Vollbeschäftigungsniveau. Er bringt diese Überlegungen in folgenden log-linearen Gleichungen zum Ausdruck:²⁸

$$\pi = \pi^e + \alpha(\hat{Y} - \hat{Y}^e) + \gamma(\log y - \log y^e) \quad (3.9)$$

²⁸ In ihren Kommentaren zu Friedmans Common Model weisen Brunner/Meltzer (1974) und Tobin (1974) jedoch darauf hin, dass die Verknüpfung dieser Gleichungen mit dem Common Modell nicht eindeutig sei.

$$\hat{y} = \hat{y}^e + (1 - \alpha)(\hat{Y} - \hat{Y}^e) - \gamma(\log y - \log y^e) \quad (3.10)$$

Die beiden Gleichungen (3.9) und (3.10) erklären zusammen mit Gleichung (3.8), wie die Änderungen der Geldmenge den Preis- und Mengeneffekte generieren. Z.B. eine Geldmengeerhöhung lässt zunächst das Nominaleinkommen über das erwartete Niveau steigen. Die dadurch bewirkte Abweichung der tatsächlichen von der erwarteten Wachstumsrate des Nominaleinkommens ($\hat{Y} - \hat{Y}^e$) wirkt auf die Inflationsrate und den Output, wobei Aufteilung der Wirkung von den Parametern α und $(1 - \alpha)$ abhängt.²⁹ Dabei bezeichnet das Verhältnis $(\log y - \log y^e)$ den Kapazitätsindex. Wenn $(\log y - \log y^e) > 0$, geht es um eine Überauslastung der vorhandenen Kapazitäten, die das reale Wachstum reduziert (Gleichung (3.10)) und die Inflation beschleunigt (Gleichung (3.9)); umgekehrt wirken Überkapazitäten dämpfend auf die Inflationsrate und beschleunigend auf das reale Wachstum.

Man kann anhand der beiden Gleichungen (3.9) und (3.10) die Lösungen von einfacher Quantitätstheorie und von traditioneller keynesianischen Ökonomie verdeutlichen. Die einfache Quantitätstheorie unterstellt, dass eine Änderung des Nominaleinkommens sich in einer Änderung des Preisniveaus niederschlägt und der Output immer auf seinem Trendniveau bleibt. Dieses Ergebnis erhält sie durch die Annahmen $\alpha = 1$ und $\gamma = \infty$. Dagegen kommt die einfache keynesianische Ökonomie zum Ergebnis, eine Änderung des Nominaleinkommens münde in eine Änderung des Outputs bei Unterbeschäftigung und in eine Änderung des Preisniveaus bei Vollbeschäftigung. Sie nimmt also an, dass gilt: $\alpha = \gamma = 0$ für $y < y^e$ und $\alpha = 1$ und $\gamma = \infty$ für $y = y^e$. Aus dieser Perspektive sieht Friedman die Gleichungen (3.9) und (3.10) als eine allgemeinere Formulierung der Teilung der Änderung des Nominaleinkommens in Preis- und Outputeffekt. Anders als die beiden anderen Lösungen spezifiziert Friedmans „third way“ nicht die Anpassungspfade von Preisniveau und Output.

Wie erklärt sich dann der Anpassungsprozess? Friedman beschreibt den Anpassungsprozess mit dem Realkasseneffekt, den wir oben dargestellt haben. Wir sparen uns die Wiederholung, aber ein Punkt ist zu unterstreichen. Im Modell Friedmans bestimmt die Änderung der Geldmenge die

²⁹ Tobin (1974: 83f.) weist darauf hin, dass die Inflationsgleichung (3.9) als eine Phillipskurve interpretiert werden kann: Die Variable $(\hat{Y} - \hat{Y}^e)$ beziehe sich auf die Änderung der Arbeitslosigkeit und die Variable $(\log y - \log y^e)$ auf deren Niveau, wobei y^e der natürlichen Arbeitslosenquote entspreche. Die langfristig vertikale Phillipskurve sei durch die erwartete Preisveränderung π^e mit dem Koeffizienten 1 gesichert.

Entwicklung des Nominaleinkommens, das ansonsten auf dem Pfad der natürlichen Arbeitslosenquote ist. Das impliziert, dass der Anpassungsprozess durch die Diskrepanz zwischen Angebot und Nachfrage der nominellen Geldmenge ausgelöst wird.

Fassen wir zusammen. Die theoretische Bedeutung der monetären Analyse Friedmans besteht in dem Versuch, die klassische Dichotomie zwischen Geld- und Realsphäre zu überwinden: „Money is all that matters for changes in nominal income and for short-run changes in real income“ (Friedman 1974: 27). Die Friedmansche monetäre Analyse bereichert damit die einfache Quantitätstheorie um die kurzfristigen Effekte des Geldes. Aber die Kernaussagen der Quantitätstheorie bleiben über den Realkasseneffekt gewahrt: Geld ist langfristig neutral und Inflation ist ausschließlich ein monetäres Phänomen. Man kann daher zu Recht davon sprechen, dass sein Interaktionsmodell im Wesentlichen die Geld- unter die Realsphäre subsumiert. Dogmengeschichtlich bedeutet dieser Umstand eine Subsumierung der keynesianischen Ökonomie unter die neoklassische Ökonomie. Seitdem bleibt die monetäre Analyse im Rahmen des Monetarismus. Die monetaristische These wurde damit (zu Unrecht) zum allgemeinen Konsens der Ökonomie erhoben.

3-3 Die um die Erwartungen erweiterte Phillipskurve

Vom Standpunkt der Neuquantitätstheorie und der natürlichen Arbeitslosenquote her hat Friedman die stabile Phillipskurve und damit die wirtschaftspolitische Ausbeutbarkeit des Trade-off zwischen Inflation und Arbeitslosigkeit in Frage gestellt. Die Einführung der adaptiven Erwartungshypothese verhilft Friedman zur Formulierung einer um die Erwartungen erweiterten Phillipskurve, die zeigt, dass sich die Phillipskurve verschiebt, wenn die Inflationserwartungen sich ändern. Dieser Umstand schließt ein, dass kurzfristig viele (und daher unstabile) Phillipskurven mit den Inflationserwartungen als Lageparameter existieren können. Hervorzuheben ist dabei, dass seine These der langfristigen Neutralität und kurzfristigen Nicht-Neutralität des Geldes in der Form einer kurzfristig links geneigten und langfristig vertikalen Phillipskurve zum Ausdruck kommt. Wir wenden uns dieser Argumentation zu.

Zunächst die Funktion des Arbeitsmarkts. Für die Friedmansche Erklärung der kurzfristigen Allokation auf dem Arbeitsmarkt sind zwei Punkte ausschlaggebend:

- Erstens ist das Arbeitsangebot eine Funktion des erwarteten Reallohns.³⁰ Formal heißt das:

$$N^s = b \left(\frac{W}{P^e} \right) = b \left[\left(\frac{W}{P} \right) \left(\frac{P}{P^e} \right) \right] \quad (3.11)$$

Die Gleichung (3.11) besagt, dass das Arbeitsangebot nicht nur von dem Reallohn, sondern auch von der Abweichung des Preisniveaus vom erwarteten Preisniveau abhängt. Die Arbeitsangebotskurve verschiebt sich mit einer Variation in P/P^e . Wenn das gegenwärtige Preisniveau das erwartete übersteigt ($P > P^e$), verschiebt sich die Arbeitsangebotskurve nach rechts und vice versa.

- Der zweite Punkt bezieht sich auf die Erwartungsbildung der Wirtschaftssubjekte. Friedman hat hier die Hypothese der adaptiven Erwartungen unterstellt, die besagt, „that anticipations are revised on the basis of the difference between the current rate of inflation and the anticipated rate (...) As is well known, this implies that the anticipated rate of inflation is an exponentially weighted average of past rates of inflation, the weights declining as one goes back in time” (Friedman 1975: 25). Formal ausgedrückt lautet sie

$$\pi_t^e = \beta \sum_{n=0}^{\infty} (1-\beta)^n \pi_{t-n-1} \quad (3.12)$$

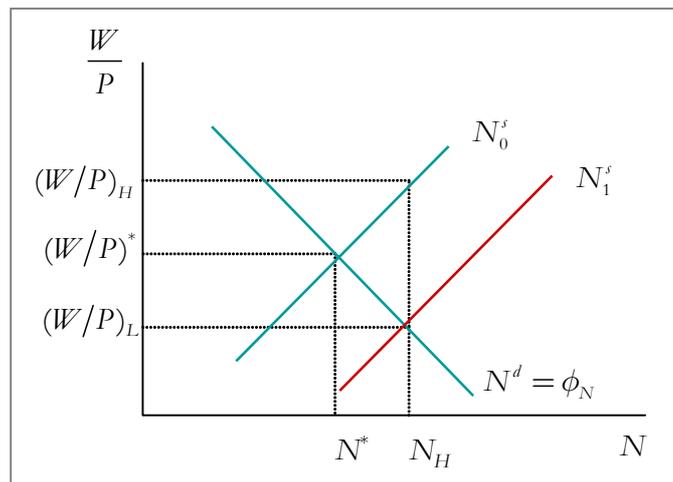
wobei der Faktor $0 < \beta < 1$ als die Anpassungsgeschwindigkeit bzw. Lernkoeffizient bezeichnet wird. In diesem „backward-looking Modell“ werden inflationäre Entwicklungen langsam wahrgenommen, sodass die Erhöhung des Preisniveaus als Veränderung der relativen Preise missgedeutet werden kann. Bezogen auf die Löhne bedeutet dies: Wenn das Preisniveau (durch eine Erhöhung der Geldmenge) erhöht wird, dann können die Arbeiter die dadurch gestiegenen Geldlöhne als Anstieg ihrer Reallöhne (miss)interpretieren, weil sie die erhöhten Geldlöhne auf Basis des vergangenen Preisniveaus bewerten werden; und vice versa. Es entsteht damit eine Geldillusion, die jedoch mit der adaptiven Anpassung der Erwartungen verschwinden wird.

Die Implikationen lassen sich anhand der Abbildung 3.2 illustrieren. Der Arbeitsmarkt ist bei der Beschäftigungsmenge N^* und dem Reallohn $(W/P)^*$ im Gleichgewicht. Nehmen wir nun einen

³⁰ Im Friedmans Modell, und in der neoklassischen Ökonomie überhaupt, wird dem Arbeitsangebot eine zentrale Rolle bei der Bestimmung von Output und Beschäftigung zugewiesen. Welche Menge an Arbeit angeboten wird, hängt hier von den Präferenzen der Arbeiter ab.

monetären Impuls an. Eine vermehrte nominale Geldmenge ist bei zunächst noch unverändertem Preisniveau mit einer Erhöhung der realen Geldmenge gleichbedeutend. Der Realkasseneffekt beginnt zu wirken. Es ergibt sich dann eine Nachfragesteigerung auf dem Gütermarkt, auf die die Unternehmen mit einer Erhöhung der Produktion zu reagieren versuchen. Entsprechend steigt die Arbeitsnachfrage an. Ein Output über dem den Arbeitsmarkt räumenden Niveau verlangt jedoch eine Beschäftigung N_H , die über dem Gleichgewichtsniveau N^* liegt. Damit die Arbeitnehmer diese Menge Arbeit anbieten können, ist ein Reallohn von $(W/P)_H$ erforderlich; andererseits müsste für die Unternehmen der Reallohn $(W/P)_L$ sein, damit sie eine solche Menge Arbeit nachfragen.

Abbildung 3.2 Der Arbeitsmarkt bei Geldillusion



Diese Differenz löst sich dadurch, dass sich die Arbeitsangebotskurve von N_0^s zu N_1^s verschiebt. Die Verschiebung findet statt, weil die Arbeiter der Geldillusion unterliegen: Sie interpretieren die durch den allgemeinen Nachfrageschub erzeugte Steigung des nominalen Lohnsatzes als eine Änderung des Reallohnsatzes. Ihre Erwartungen der künftigen Preise bleiben zu diesem Zeitpunkt noch unverändert, und sie bewerten das Nominallohnangebot auf Basis dieser Erwartungen. Sie glauben daher, dass der Reallohn höher ist, als er wirklich ist. Sie bieten folglich eine höhere Arbeitsmenge an als beim tatsächlichen Reallohn angeboten wäre. Im Gegensatz zu den Arbeitnehmern sollen jedoch die Unternehmen die Entwicklung des Preisniveaus präzise einschätzen. Sie gehen von sinkenden Reallöhnen aus und fragen die Beschäftigung N_H nach. Eine unerwartete Inflation führt zu einer falschen Einschätzung der Reallohnentwicklung seitens der Arbeitnehmer

und damit zu einem höheren Arbeitsangebot. Die positive Relation von Geldmenge und Beschäftigung in der Phillipskurve funktioniert deshalb, weil eine Geldillusion seitens der Arbeitnehmer besteht.

„Indeed, the simultaneous fall ex post in real wages to employers and rise ex ante in real wages to employees is what enabled employment to increase“ (Friedman 1968: 10).

So weit die Logik der Phillipskurve, wie sie von Friedman reformuliert wurde. Friedmans Kritik an der Phillipskurve besteht darin, dass sie konstante Preiserwartungen postuliert habe. Dadurch wäre die Argumentation der Phillipskurve im Grunde eine statische Analyse und hätte den Lernprozess der Wirtschaftssubjekte ignoriert.

Im Friedmans Modell sind die Erwartungen nicht konstant, sondern bilden sich adaptiv. Das heißt, die Arbeiter werden mit der Zeit die tatsächliche Inflationsrate erkennen und von der Geldillusion frei, indem sie ihren Erwartungsfehler korrigieren. Die Arbeitsangebotskurve wird daraufhin nach links verschoben, so dass der Arbeitsmarkt zur alten Position mit einer Kombination von $(W/P)^*$ und N^* zurückkehrt. Der kurzfristige Effekt des Geldes auf die Beschäftigung verschwindet und die Neutralität des Geldes bleibt gewahrt.

„You can fool all of the people some of the time, you can fool some of the people all of the time, but you can't fool all of the people all of the time“ (Friedman 1975: 21).

Vor diesem Hintergrund unterscheidet Friedman zwischen einer langfristigen und einer kurzfristigen Phillipskurve: Es kann aufgrund der Geldillusion eine inverse Beziehung zwischen Inflationsrate und Arbeitslosigkeit entstehen, aber langfristig setzt sich eine natürliche Arbeitslosenquote durch, die unabhängig von monetären Faktoren ist. Die Inkorporierung der Inflationserwartungen in die Phillipskurve erbringt das Ergebnis, dass für unterschiedliche Inflationserwartungen unterschiedliche kurzfristige Phillipskurven existieren. Andererseits verläuft die Phillipskurve langfristig vertikal bei der natürlichen Arbeitslosenquote, die als einziges Gleichgewicht der Ökonomie gilt. Langfristig werden die Erwartungsfehler ausgeglichen und entsprechend kehrt die Ökonomie ins Gleichgewicht zurück. Wir führen diese Zusammenhänge weiter aus.

Folgt man dem Friedmanschen Postulat, die ökonomisch relevante Größe für die Beschäftigung sei nicht der Nominal-, sondern der Reallohn, kann man die Gleichung (2.1) ersetzen durch

$$(\hat{w}_t - \pi_t) = f(u_t) \quad (3.13)$$

Da die tatsächliche Preisentwicklung zum Zeitpunkt t noch unbekannt ist, tritt an die Stelle von π_t die erwartete Inflationsrate π_t^e . Folglich lässt sich die Gleichung (3.13) umformulieren als

$$\hat{w}_t = \pi_t^e + f(u_t) \quad (3.14)$$

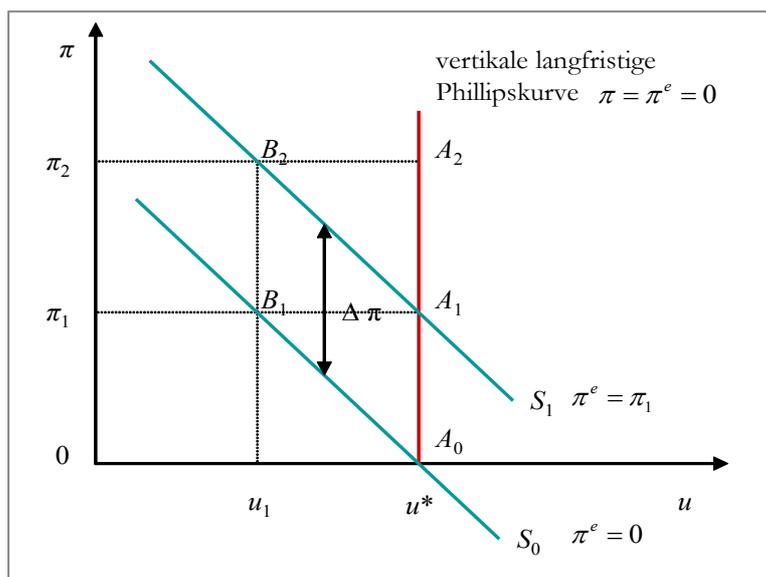
Der Nominallohn hängt nun sowohl von dem Beschäftigungsniveau als auch von den Inflationserwartungen ab. Unter Annahme einer konstanten Produktivitätsentwicklung wird die Gleichung (3.14) zu

$$\pi_t = \pi_t^e + f(u_t) \quad \text{bzw.} \quad (3.15)$$

$$\pi_t - \pi_t^e = -\alpha(u_t - u^*) \quad (3.16)$$

Die Gleichung (3.16) ist eine spezifizierte Form von Gleichung (3.15), wobei u^* die Vollbeschäftigungsarbeitslosenquote, also die NRU, darstellt. Der Koeffizient $\alpha > 0$ bezeichnet die Neigung der Phillipskurve und bestimmt den Trade-off zwischen Inflation und Arbeitslosigkeit. Man sieht, dass die erwartete Inflationsrate als zusätzliche Variable auf der rechten Seite der Gleichung erscheint. Von daher bezeichnet man die Gleichung (3.15) als die „expectations augmented Phillips curve“. Die erwartete Inflationsrate fungiert hier als Lageparameter der (kurzfristigen) Phillipskurve, d.h. die Phillipskurve verschiebt sich, wenn die erwartete Inflationsrate sich verändert.

Abbildung 3.3 Die um die Erwartungen erweiterte Phillipskurve



Stellen wir diese Überlegungen anhand der Abbildung 3.2 dar. Beginnen wir mit einer stabilen Position A_0 mit der NRU u^* und $\pi_t = \pi_t^e = 0$. Nehmen wir an, dass eine expansive Geldpolitik eingesetzt wird, um die Arbeitslosenquote von u^* auf u_1 zu reduzieren.³¹ Die Geldmengenerhöhung löst eine unerwartete Inflation ($\pi_t > \pi_t^e = \pi_{t-1}$) aus und übt einen positiven Effekt auf die Beschäftigung aus, so dass die Arbeitslosenquote unter dem natürlichen Niveau liegt ($u_t = u_1 < u^*$). Die Erhöhung des Outputs und damit der Beschäftigung über das natürliche Niveau ist deshalb möglich, weil die Arbeiter ihre Inflationserwartungen adaptiv bilden. Sie betrachten die Erhöhung der Nominallöhne auf Basis der Inflationsrate der Vorperiode, so dass die Nominallohnerhöhung als Erhöhung des Reallohnes missinterpretiert und mehr Arbeit angeboten wird. Das heißt, sie unterliegen einer temporären Geldillusion. Die Ökonomie bewegt sich damit entlang der kurzfristigen Phillipskurve S_0 zum Punkt B_1 .

Nun beginnen die Arbeiter, ihre Inflationserwartungen an die tatsächliche Inflationsrate anzupassen. Entsprechend wird eine Erhöhung des Nominallohnes um die erwartete Inflationsrate gefordert. Somit verschiebt sich die kurzfristige Phillipskurve von S_0 zu S_1 , wenn $\pi_t^e = \pi_1$ wird. Da der Reallohn nun gestiegen ist, drosseln Unternehmen ihre Arbeitsnachfrage. Die Beschäftigung kehrt auf das natürliche Niveau zurück. Der Punkt A_1 markiert dann eine neue Gleichgewichtsposition, aber mit einer gestiegenen Inflationsrate, die dem Geldmengenwachstum korrespondiert.

Hält die Geldpolitik an dem Ziel u_1 fest, dann muss das Geldmengenwachstum weiter erhöht werden. Ein Expansionseffekt auf Output und Beschäftigung, der nun auf Basis der Inflationserwartungen π_1 ausgelöst wird, ist jedoch temporär und instabil. Im Punkt B_2 findet erneut der soeben geschilderter Anpassungsprozess statt. Die erwartete Inflationsrate wird $\pi_t^e = \pi_2$, so dass sich die Phillipskurve sich um die gestiegene Inflationsrate nach oben verschiebt ($S_1 \rightarrow S_2$). Die Ökonomie wird auf das neue Gleichgewicht A_2 geführt. Kombiniert man die Punkte A_0, A_1, A_2, \dots , dann ergibt sich eine langfristige Phillipskurve, die bei der natürlichen Arbeitslosenquote vertikal verläuft.

Dieser Sachverhalt impliziert, dass es keinen Trade-off zwischen Inflation und Beschäftigung gibt, sobald die Inflationsrate vollständig antizipiert wird. Die monetären Impulse verpuffen, ohne kurz-

³¹ Diese Argumentation ist symmetrisch. Das heißt, für den Deflationsprozess gilt die umgekehrte Logik. Dieser Sachverhalt wird in Literatur im Zusammenhang mit der Disinflation diskutiert. Wir verzichten jedoch auf deren detaillierte Darstellung.

fristige Realeffekte, vollständig in einer gestiegenen Inflationsrate. Auch wenn man Erwartungsfehler zulässt, wird der Versuch, mithilfe der expansiven Geldpolitik die Beschäftigung über dem natürlichen Niveau zu halten, in eine akzelerierende Inflation münden. Er erfordert eine kontinuierlich steigende Wachstumsrate der Geldmenge („accelerationist hypothesis“). Aus diesen Überlegungen hat Friedman die Schlussfolgerung gezogen:

„There is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off. The temporary trade-off comes not from inflation per se, but from unanticipated inflation, which generally means, from a rising rate of inflation” (Friedman 1968: 11).

3-4 Wirtschaftspolitische Implikationen: Das neoklassische Assignment

Das Modell Friedmans lässt sich als Versuch charakterisieren, das klassische Neutralitätspostulat mit kurzfristigen Realeffekten des Geldes zu versöhnen – was in der Unterscheidung zwischen kurzfristig links geneigter und langfristig vertikaler Phillipskurve zum Ausdruck kommt. Das Modell Friedmans weist jedoch eine dichotomische Struktur auf: Die lange Frist funktioniert nach dem neoklassischen Prinzip plus der Quantitätstheorie; in der kurzen Frist gilt die keynesianische Theorie in dem Sinne, dass die (unerwarteten) Geldmengenveränderungen aufgrund der Geldillusion Wirkungen auf Output und Beschäftigung zeitigen. Das Modell Friedmans transformiert damit die klassische Dichotomie zwischen Geld- und Gütersphäre in die Dichotomie zwischen der langfristigen Neutralität und kurzfristigen Nichtneutralität des Geldes – ein Prinzip, das, zu Unrecht, zum allgemeinen Konsensus der Wirtschaftstheorie erhoben wird.³²

Aus dieser Perspektive werden zwei wirtschaftspolitische Ansätze hergeleitet, die immer noch die wirtschaftspolitische Diskussion dominieren: Preisstabilität durch potentialorientierte Geldmengengerel und Beschäftigung durch angebotsorientierte Arbeitsmarktpolitik – das ‚neoklassische Assignment‘ ist geboren.

³² Dass die langfristige Gültigkeit der Quantitätstheorie breiten Konsens genießt, zeigt sich daran, dass sich die Neu-Keynesianer bezogen auf diese Frage als Monetaristen bezeichnen. “When it comes to understanding inflation over the longer term, economists typically emphasize just one factor: growth of money supply. The reason for this emphasis is that no other factor is likely to lead to persistent increases in the price level” (Romer 2001: 470).

3-4-1 Preisstabilität und Geldmengenregel

Der neoklassische Monetarismus versteht die Geldpolitik als Geldmengensteuerung, deren primäres Ziel die Gewährleistung der Preisstabilität ist.³³

Die exklusive Ausrichtung auf die Preisstabilität resultierte einerseits aus der historischen Erfahrung der inflatorischen Entwicklung in den 1970er Jahren. Sie reflektierte andererseits einen Regimewechsel, der vom Paradigmawechsel von keynesianischer Ökonomie zum neoklassischen Monetarismus begleitet wurde. Während der traditionelle Keynesianismus auf Basis der Phillipskurve für eine Politik leichten Geldes zur Vollbeschäftigungsgarantie eingetreten war, hat der neoklassische Monetarismus hervorgehoben, dass das Reiten auf der Phillipskurve letztlich ohne Wirkungen auf Output und Beschäftigung in eine (akzelerierende) Inflation mündet. Da die Inflation jedoch Wohlfahrtskosten für das ökonomische System beinhaltet, wurde die Preisstabilität als vorrangiges Ziel der Geldpolitik vorgeschrieben.³⁴

Die Geldpolitik soll daher über die Geldmengensteuerung die Preisstabilität gewährleisten. Zu diesem Zweck hat Friedman eine Geldmengenregel vorgeschlagen:

„My own prescription is still that the monetary authorities go all the way in avoiding such swing by adopting publicly the policy of achieving a steady rate of growth in a specified monetary total” (Friedman 1968: 16).

³³ Diese monetaristische Grundorientierung fand ihre politische Anwendung vor allem in der Geldpolitik der deutschen Bundesbank von 1975 bis 1998 sowie im monetaristischen Experiment der Federal Reserve Bank unter Paul Volcker von 1978 bis 1982. Neuerdings spielt diese Geldmengensteuerung eine prominente Rolle bei der geldpolitischen Strategie der EZB, indem sie erste Säule der Zwei-Säulen-Strategie der EZB bildet. Allerdings weist eine Reihe von Ökonomen darauf hin, dass die Bundesbank trotz ihres konzeptionellen Bekenntnisses zum „monetary targeting“ opportunistisch und diskretionär gehandelt habe (von Hagen 1995). Clarida/Gertler (1998) stellen auch heraus, dass die Zinspolitik der Bundesbank, analog zur Taylor-Regel (Taylor 1993), auf die Abweichungen des realen Wachstums und der Inflation von ihrem Zielniveau reagiert hat.

³⁴ Unbestritten ist, dass die Aufgabe der Geldpolitik die Sicherung des Geldwertes ist. Es ist jedoch eine offene Frage, welche Rolle der Geldpolitik darüber hinaus zufällt. Zur Debatte steht vor allem, ob die Geldpolitik neben der Preisstabilität eine Stabilisierung der Beschäftigung ins Kalkül zu nehmen hat. Während die Federal Reserve Bank an diese Frage flexibel herangeht, distanziert sich die EZB von einer Beschäftigungsstabilisierung als Zielsetzung (Issing 2001: 68). Der EG-Vertrag, der die Ziele und Aufgaben der EZB vorschreibt, legt fest, dass das vorrangige Ziel der EZB die Sicherung der Preisstabilität ist und die EZB, soweit dieses Ziel nicht beeinträchtigt wird, die anderen wirtschaftspolitischen Ziele unterstützt (vgl. Artikel 105 (1) EG-Vertrag). Wegen dieser Fixierung auf die Preisstabilität ist die Geldpolitik der EZB oft in die Kritik geraten.

Nach diesem Vorschlag soll sich der Zuwachs der Geldmenge – verstanden als Geldangebot – an der langfristigen Wachstumsrate des Produktionspotentials, korrigiert um die Veränderung der Umlaufgeschwindigkeit, ausrichten. Der Referenzwert der potentialorientierten Geldmengenwachstumsrate leitet sich dabei aus der Quantitätsgleichung ab.³⁵

$$\hat{M}^{Ziel} = \pi^{Norm} + \hat{y} - \hat{v}$$

Der Akzent liegt dabei nicht darauf, wie die Geldmenge definiert wird und welche Rate vorgesehen ist, sondern auf der Verstetigung des Geldmengenwachstums.³⁶ Eine solche Regel soll durch Unterbindung der diskretionären Handlungen der Zentralbank das Vertrauen in den geldpolitischen Kurs stärken und damit die Glaubwürdigkeit erhöhen. Die Konstanz der monetären Bedingungen macht demnach die Geldpolitik voraussehbar und stabilisiert damit die Inflationserwartungen.

Friedman leitet die Anforderungen an die Geldmengenregel aus der Grundannahme ab, dass die Realsphäre immanent stabil und sich selbst regulierend ist; eine inadäquate Geldpolitik löst unnötige Schwankungen der Realsphäre aus, die dann diese „Schocks“ zu verarbeiten hat. Geld ist eine Quelle der Störungen, die durch eine Geldmengenregel ausgeschaltet werden soll.³⁷ Darüber hinaus sollte beachtet werden, dass die Geldpolitik nicht imstande sei, die Realgröße wie Output und Beschäftigung sowie Realzins auf ein vorausbestimmtes Niveau festzulegen. Der Grund liegt, so Friedman, in unüberwindbaren Unsicherheiten und unvollkommenem Wissen über die Funktionsweise der Ökonomie. So können z.B. die langen und variablen Wirkungszögerungen des monetären Transmissionsmechanismus dazu führen, dass aktive, antizyklische Wirtschaftspolitik die ökonomischen Schwankungen verstärkt statt dämpft.

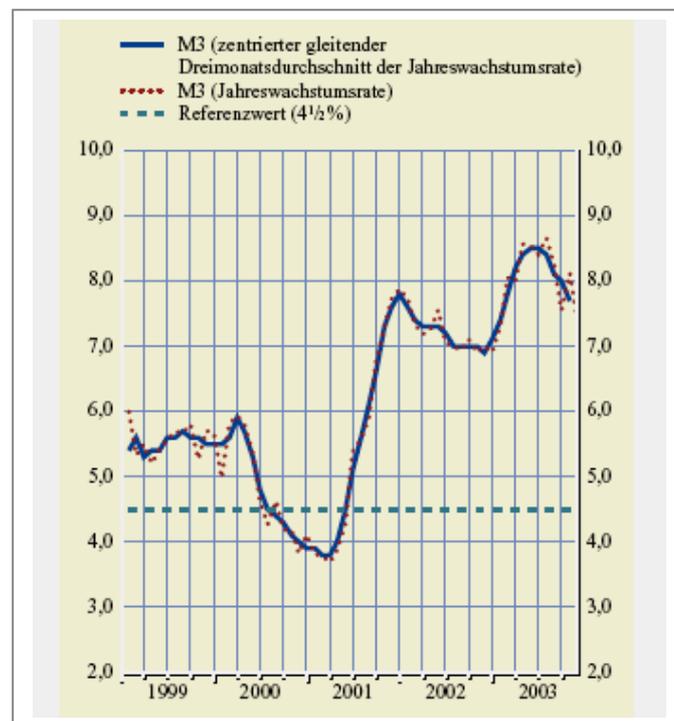
³⁵ Dem monetaristischen Prinzip folgend und die Geldpolitik der deutschen Bundesbank fortführend, gibt die EZB im Rahmen ihrer Zwei-Säule Strategie der Geldpolitik als Referenzwert die Entwicklung des Geldmengenaggregats M3 an. Sie zielt auf ein Wachstum des Aggregats vom 4.5% pro Jahr ab. Diese Zielgröße ergibt sich aus ihrer Definition der Preisstabilität (einer Inflationsrate zwischen 0% und 2%), einer angenommenen Wachstumsrate des Produktionspotentials von 2% bis 2.5% sowie aus einem trendmäßigen Rückgang der Umlaufgeschwindigkeit des Geldes von 0.5% bis 1% (EZB 1999, 2000b, 2003).

³⁶ Der neoklassische Monetarismus zieht als Steuerungsgröße das Geldmengeaggregat M3 vor, das Geld im weitesten Sinne erfasst. Ein Grund dafür ist, dass es im Vergleich zu M1 und M2 bei Zinsänderungen gegen Effekte der Portfolioumschichtung resistent sei. In dieser Praxis reflektiert sich, dass die Quantitätstheorie mit dem Homogenitätspostulat die Zinsabhängigkeit der Geldnachfrage eliminiert. Aber Heine/Herr (2001: 54ff.) haben gezeigt, dass nicht nur die Geldmengenaggregate M1 und M2 sondern auch M3 erheblichen Schwankungen unterliegen.

³⁷ „It (Geld – PKS) throws a monkey wrench into the operation of all the other machines“ (Friedman 1968: 12)

Das Konzept der Geldmengenregel ist jedoch auf zwei fragwürdigen Annahmen angewiesen. Erstens wird postuliert, dass die Zentralbank in der Lage ist, die Geldmenge hinreichend zu kontrollieren, so dass aus Sicht des privaten Sektors die Geldmenge exogen gesetzt wird. Die Annahme exogenen Geldangebots zwingt zweitens, eine stabile Geldnachfrage zu postulieren. Eine stabile Geldnachfrage – sie kommt einer stabilen Umlaufgeschwindigkeit des Geldes gleich – ist für die Quantitätstheorie so fundamental, dass die Steuerung des Preisniveaus durch die Geldmenge nur unter dieser Bedingung möglich ist. Ansonsten wird der Zusammenhang zwischen Geldmenge und Preisniveau zerstört. Die beiden Annahmen implizieren, dass die monetären Schocks im Prinzip vom Geldangebot herrühren und daher für die Inflation (und Deflation) der Schurke im Stück die Zentralbank bzw. ihre diskretionären Handlungen ist.

Abbildung 3.4 M3-Wachstum und der Referenzwert im Euro-Raum



Quelle: EZB Monatsbericht Januar 2004.

Aber wie selbst die Bundesbank (1995: 92), die als Paradefall der Geldmengensteuerung gilt, zugegeben hat, entsteht in funktionierenden Geldwirtschaften das Zentralbankgeld endogen. Dies impliziert, „dass eine Zentralbank den Wirtschaftssubjekten ihr Geld nicht aufzwingen, sondern lediglich ihre Zinssätze senken kann“ (Heine/Herr 2001: 25). Dass die Zentralbank keine monetäre Autorität, sondern eine Marktteilnehmerin ist und daher schwerlich ihr Geldmengeziel durchsetzen

kann, zeigt beredt der Umstand, dass die Bundesbank regelmäßig ihre Geldmengenziele verfehlt (Tomann 1997: 94, Tabelle 7.3).³⁸ Ebenso zeigt die Abbildung 3.4, dass die Wachstumsrate von M3 im Euro-Raum deutlich über dem Referenzwert 4.5% liegt – aber ohne inflationäre Effekte. All diese Fakten sprechen dafür, dass das sog. „money-supply targeting“ keine seriöse Option der Geldpolitik ist (Blinder 1998: 27ff.), und erklären, warum die Zentralbanken in fast allen Industrieländern derzeit von jener Geldmengesteuerung zum zinsgesteuerten Inflations-Targeting übergegangen sind.

Des weitern ist zu bemerken, dass die Geldmengensteuerung („monetary targeting“) im Prinzip restriktiv agiert.³⁹ Denn die strikte Fixierung auf die Geldmenge, die von den Monetaristen als eigentliche Ursache der Inflation identifiziert wird, hat zur Folge, dass die Zentralbank die inflationären Gefahren überbewerten und somit beschäftigungspolitische Chancen verspielen würde. Heine/Herr (2001: 47) zufolge ist eine solche Geldpolitik in doppelter Hinsicht schädlich: „Zum einen akzeptiert sie ohne geldpolitische Zwänge ein unnötig hohes Ausmaß an Arbeitslosigkeit und untergräbt so das Vertrauen in die Notwendigkeit einer unabhängigen Zentralbank und zum anderen honoriert sie nicht, dass sie durch die Tarifpolitik entlastet wird.“

3-4-2 Angebotsorientierte Arbeitsmarktpolitik.

Eine zentrale Aussage des neoklassischen Monetarismus ist, dass der Arbeitsmarkt durch die sich selbst regulierenden Kräfte zum (Vollbeschäftigungs-)Gleichgewicht tendiert, das mit der natürlichen Arbeitslosenquote assoziiert ist, die mikroökonomisch und realökonomisch determiniert wird. Dabei wird das Gleichgewicht am Arbeitsmarkt durch den Realkasseneffekt sichergestellt. Wenn z.B. eine temporäre nachfragebedingte Arbeitslosigkeit auftritt, findet eine Geldlohn-Preis-

³⁸ Estrella/Mishkin (1996) zeigten auf, dass eine stabile Beziehung zwischen monetären Aggregaten und Preisstabilisierung als Endziel der Geldpolitik schwer zu bestätigen sei und daher die monetären Aggregate als Zwischenzielgröße nicht geeignet seien.

³⁹ Solow (2000: 16) gibt einen instruktiven Hinweis auf die restriktiven Wirkungen der monetaristischen Geldmengensteuerung: „If the Bundesbank were sitting in Washington, the American prosperity would have been cut off by higher interest rates at least two years ago, and more likely four years ago. The U.S. would have lost two years of upswing, maybe four. The unemployment rate would be at least one and a half points higher than it is now. As a reward, the Consumer Price Index which actually rose by 1.5 percent in the past year would have risen perhaps by only 1.2 percent. This does not seem like a bargain.“

Deflation statt, welche zu höherer Realkassenhaltung beim Publikum und damit zu größerer Güternachfrage führen soll. Als Folge stellt sich das Vollbeschäftigungsgleichgewicht wieder her.

Die Implikation, die daraus gefolgert werden kann, ist, dass die Beschäftigung nicht durch eine Makropolitik der Nachfragesteuerung, sondern allein durch die auf die Angebotsseite ausgerichteten strukturellen Maßnahmen verbessert werden kann. Mit der Durchsetzung einer solchen monetaristischen Logik einhergehend – sprich: der „monetaristischen Konterrevolution“ –, tritt die angebotsorientierte Politik auf den Plan, die einen fast monopolartigen Status bei der derzeitigen wirtschaftspolitischen Diskussion besitzt.

Nach der monetaristischen Logik, die der Angebotspolitik die theoretische Legitimation liefert, ist der Reallohn der bestimmende Faktor für die Arbeitslosigkeit. Man kann diesen Zusammenhang anhand der Quantitätsgleichung verdeutlichen. Die Quantitätsgleichung (3.1) kann in einer logarithmierten Form umgeschrieben werden: $\hat{y} = \hat{m} - \pi + \hat{v}$. Da der Zuwachs der Geldmenge exogen und die Änderungsrate der Umlaufgeschwindigkeit konstant bzw. eine exogene Rate ist, lässt sich diese Gleichung umformulieren als

$$\hat{y} = \hat{m} - \pi$$

Die mark-up-Beziehung gibt an $\pi = \hat{w} - \hat{\lambda}$, wenn wir den mark-up konstant halten (vgl. Gleichung (2.6) im Kapitel 2). Aus beiden Gleichung ergibt sich

$$\hat{y} = \hat{m} - \hat{w} + \hat{\lambda}$$

Die Beschäftigung ist nach der Definition $H = y/\lambda$ bzw. $\hat{h} = \hat{y} - \hat{\lambda}$.⁴⁰ Dies eingesetzt ergibt sich:

$$\hat{h} = \hat{m} - \hat{w}$$

Die Lohnentwicklung bestimmt die Entwicklung der Beschäftigung. Die so skizzierten Zusammenhänge zeigen auf, dass die Deduktion der kausalen Abhängigkeit der Beschäftigung von der Höhe der Nominallohne die quantitätstheoretischen Annahmen, nämlich die exogene Vorgabe der Wachstumsrate der Geldmenge durch die Zentralbank, voraussetzt.

⁴⁰ $N^d = N^d \cdot y/y = y/\lambda$ mit der Arbeitsproduktivität $\lambda = y/N$.

Der neoklassische Monetarismus und die von ihm inspirierte Angebotspolitik propagierten auf dieser Grundlage die Lohnzurückhaltung als entscheidendes Instrument zur Bekämpfung der Arbeitslosigkeit. Demnach soll die Lohnzurückhaltung eine Verbesserung der Erlös-Kosten-Relation bewirken, so dass die lohnpolitischen Impulse über die Gewinne und Investitionen auf die Beschäftigung transformiert werden. Die Frage in diesem Kontext kann nur lauten: „Wer also die Lohnhöhe als Ursache der Arbeitslosigkeit beklagt und sich damit für eine Lohnkorrektur ausspricht, muss sagen, welche die Lohnhöhe bestimmenden exogenen Variablen in welcher Weise und von wem beeinflusst werden sollen, damit sich die für ihn als adäquat angesehene Lohnhöhe als Resultat ergibt“ (Franz 1992: 10).

Aus diesem Blickwinkel diagnostiziert die Angebotspolitik, dass die hohe und dauernde Arbeitslosigkeit in den EU-Ländern in den starren Strukturen des Arbeitsmarktes begründet liege.⁴¹ Diese hätten eine teure und inflexible Arbeit zur Folge, die eine Ausweitung der Produktion behindere. Das Stichwort lautet: Eurosclerose. Demnach hätten institutionelle Verkrustungen ein Ausmaß erreicht, das eine vollbeschäftigungskonforme Selbstregulierung des Arbeitsmarktes unmöglich mache. Das Zielobjekt der Eurosclerose-Diagnose ist freilich neben den Gewerkschaften der Sozialstaat. Die Therapie ist eindeutig: Die kosten- und anreizbezogenen Verzerrungen der Angebotsbedingungen am Arbeitsmarkt müssen beseitigt werden, damit der Unternehmergeist ermutigt wird und, als Folge, die Nachfrage nach Arbeit steigt. Der Code-Word heißt: Flexibilität mit dem Ziel der Absenkung des Lohnniveaus. Ohne dieser Argumentation weiter zu folgen, mag der Hinweis genügen, „dass (...) hinter den Forderungen nach Flexibilität unvermeidlich auch Verteilungsansprüche stecken“ (Ganßmann 2000b: 91). Die Vergleiche der Arbeitsmarktdaten von OECD-Ländern (Ganßmann 2000b: 95ff.) zeigen ferner, dass die Gleichung: ausgebauter Sozialstaat = unflexibler Arbeitsmarkt = hohe Arbeitslosigkeit nicht aufgeht.

„I think this (diese ausschließliche Fokussierung auf Flexibilität des Arbeitsmarktes – PKS) is in part an intellectual failure. Sometimes it is hard to escape the thought that the single-minded focus on the labor market stems from the naive belief that unemployment *must* be a defect in the labor market, as if the hole in a flat tire must always be on the bottom, because that is where the tire is flat“ (Solow 2000: 5).

Fassen wir zusammen. Der neoklassische Monetarismus, auf den sich die Angebotspolitik stützt, reduziert das Beschäftigungsproblem auf ein mikroökonomisches Allokationsproblem: Die Arbeitslosigkeit ist ein reales Phänomen, das die Makropolitik nicht auffangen kann; und sie ist ein mikroökonomisches Problem, das durch einzelwirtschaftliches Handeln wieder beseitigt wird.

⁴¹ Siehe etwa den Titel von Siebert (1997) : „Labor Market Rigidities: At the Root of Unemployment in Europe“.